AD-A019 047

PROSODIC AIDS TO SPEECH RECOGNITION: VII. EXPERIMENTS ON DETECTING AND LOCATING PHRASE BOUNDARIES

Wayne A. Lea

Sperry Univac

Prepared for:

Advanced Research Projects Agency

14 November 1975

SPERRY UNIVAC
DEFENSE SYSTEMS

PROSODIC AIDS TO SPEECH RECOGNITION:

VII. EXPERIMENTS ON DETECTING AND LOCATING PHRASE BOUNDARIES

by

Wayne A. Lea

Defense Systems Division
St. Paul, Minnesota
(612) 456-2434

Final Technical Report Submitted To:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Director, IPTO

14 November 1975

Report No. PX 11534

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Univac Defense Systems Division P.O. Box 3525 St. Paul, Minn. 55165	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Prosodic Aids to Speech Recognition: VII. Experiments on Phrase Boundary Detection and Location

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Technical Report: 1 September 1974 – 31 August 1975

5. AUTHOR(S) (First name, middle initial, last name)

Wayne A. Lea

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
14 November 1975	52	35

8a. CONTRACT OR GRANT NO. DAHC 15-73-C-0310	9a. ORIGINATOR'S REPORT NUMBER(S) Univac Report No. PX 11534
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None
d.	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency 9400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Computer programs for detecting syntactic boundaries (BOUND3) and locating stressed syllables (STRESS) have been supplied to ARPA contractors and incorporated into speech recognition facilities. Work is just beginning on how to use syntactic boundaries in parsing procedures. Experiments were conducted on various timing cues that correlate with phonological and syntactic phrase boundaries, showing that 91% of the phonological phrase boundaries that were perceived by listeners who heard spectrally inverted speech could be detected from lengthened vowels and sonorants in phrase-final positions. Also, 95% of these perceived boundaries were evidenced by long time intervals between stressed syllables. The interstress interval also provided a good measure of rate of speech, that correlated with error rates in automatic phonetic classification schemes. Analysis of fundamental frequency contours in 159 sentences with minimal contrasts in linguistic structure showed that fundamental frequency rises invariably, just before the first stressed syllable in a constitutent. Among the most robustly marked constituent boundaries is that between the verb and the object noun phrase of a sentence. A number of other specific hypotheses relating fundamental frequency contours to linguistic structures were also tested. Glottal stops were found to occur almost exclusively before stressed vowels in word-initial position. A glottal stop is usually preceded by a rising fundamental frequency contour, and followed by other large variations in fundamental frequency. The stop may (but need not necessarily) exhibit a period of unvoicing. They were found to be more likely to occur at constituent boundaries than within major constituents. Further work is planned, to test prosodic patterns associated with various sentence types and specific syntactic structures such as subordination, coordination, etc. Applications to ARPA speech understanding systems are to be investigated.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

18

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Recognition						
Speech Analysis						
Linguistic Stress						
Prosodies						
Prosodic Feature Extraction						
Syntactic Boundary Detection						
Stressed Syllable Location						
Rhythm						
Disjuncture						
Durations						
Phonological Phrase Boundaries						

SPERRY UNIVAC
DEFENSE SYSTEMS

PROSODIC AIDS TO SPEECH RECOGNITION:

VII. EXPERIMENTS ON DETECTING AND LOCATING PHRASE BOUNDARIES

by

Wayne A. Lea

Defense Systems Division
St. Paul, Minnesota
(612) 456-2434

Final Technical Report Submitted To:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Director, IPTO

14 November 1975

Report No. PX 11534

A

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the author, and should not be interpreted as necessarily representing the official policies, either expressed, or implied, of the Advanced Research Projects Agency or the U.S. Government.

1 b

PREFACE

This is the seventh in a series of reports on Prosodic Aids to Speech Recognition. The previous reports appeared as follows:

I.	Basic Algorithms and Stress Studies	1 October 1972	PX 7940
II.	Syntactic Segmentation and Stressed Syllable Location	15 April 1973	PX 10232
III.	Relationships Between Stress and Phonemic Recognition Results	21 September 1973	PX 10430
IV.	A General Strategy for Prosodically-Guided Speech Understanding	29 March 1974	PX 10791
V.	A Summary of Results to Date	31 October 1974	PX 11087
VI.	Timing Cues to Linguistic Structure and Improved Computer Programs for Prosodic Analysis	31 March 1975	PX 11239

This research was supported by the Advanced Research Projects Agency of the Department of Defense, under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

SUMMARY

During this contract year, Sperry Univac supplied ARPA Speech Understanding Research contractors with two computer programs for prosodic analysis. One detects about 90% of the boundaries between major syntactic constituents, using fall-rise patterns in fundamental frequency contours as acoustic cues to boundaries. A measure of the confidence associated with each boundary is also supplied by that program. A second program locates stressed syllables from local rises in fundamental frequency and long high-energy syllabic nuclei. This program has been shown to locate 89% of the syllables perceived as stressed by a panel of listeners.

These two programs have been made available over the ARPANET, and have been integrated into speech analysis systems at Sperry Univac and at Bolt Beranek and Newman (BBN). Work is just beginning on how to relate syntactic boundary detections to parsing procedures within the BBN system.

Besides providing prosodic analysis tools suitable for use in speech understanding systems, Sperry Univac is concerned with conducting a series of experiments on how prosodic features (i.e., fundamental frequency contours, speech energy measures, durations of linguistic units, rhythm, and pauses) provide cues to linguistic structures. In previous years, we have shown that stressed syllables provide "islands of reliability" in which carefully articulated speech is most readily decoded by machine. We found that listener's perceptions of stress patterns provide a reliable standard for determining which syllables are indeed stressed and thus should be located by any algorithm for stressed syllable location. An algorithm was developed for locating stresses at long, high-energy nuclei near the initial increase in fundamental frequency in a syntactic constituent and near subsequent rises above an "archetype contour". This algorithm, applied by hand, was shown to yield better performance in stressed syllable location than simpler procedures for locating stressed syllables from durations of syllabic nuclei alone or occurrences of increasing fundamental frequency alone.

This year, three experiments were conducted on timing cues to linguistic structure. One experiment showed that large increases in the durations of vowels and some consonants occur in phrase-final positions, so that 91% of those boundaries that were perceived when listeners heard only prosodic information (via spectrally inverted speech) were detected by groups of syllables containing segments whose lengths were 20% or more

above their median lengths. Another experiment demonstrated that one can detect major phrase boundaries from timing of prosodic features alone. Long time intervals between stressed vowels accompanied 95% of the perceived boundaries between phonological phrases. In a third experiment, the duration of the interval between two stressed syllables was found to inversely correlate with the percentage of phones that were erroneously categorized by various available methods for automatic phonetic categorization. That is, the shorter the interstress interval, the more likely that phonetic categorization errors would occur in that region of the speech. Thus, the interstress interval is a good measure of rate of speech, which might be used to predict when phonological perversions occur that phonological rules will have to handle.

These three experiments show further ways in which stressed syllables and other prosodic information could be used to determine important aspects of syntactic and phonological structure. However, as with our previous experiments on speech texts of uncontrolled structure, there are instances in the data for which it is difficult to tell which aspects of complex structures and interacting influences cause any particular prosodic patterns, such as any failure to have some syntactic boundaries marked by fall-rise patterns in the fundamental frequency contour. Carefully controlled experiments, with "near-minimal-pairs" of sentences, which are identical except for one difference in linguistic structure (or a few isolatable differences in structure), can provide precise explanations for each success or failure in determining linguistic structures from prosodic patterns. Near the end of our previous contract, a data base of 1100 sentences was designed to permit such carefully controlled experiments.

The data base includes subsets of sentences which are designed to explicitly test the prosodic effects of sentence type, contrastive syntactic bracketing, subordination, coordinate phrases and clauses, syntactic categories (such as pronouns, verbals, compound nouns, etc.), movement of stress within phrases, etc. During this contract, the 1100 sentences were recorded by three male talkers. This required a careful ordering of the sentences, to prevent pronunciations of each structure from being influenced by similar or minimally contrastive sentences that were spoken just before that sentence. By avoiding sequences of sentences that were minimally contrastive, and by interspersing sentences of different phonetic structures and various complexities, we accomplished fairly natural pronunciations. The sentences were later to be rearranged into a number of subsets, where each subset tests a particular set of prosodic or phonetic hypotheses.

Recording the sentences involved some novel procedures, including making a set of 1100 35mm slides, and projecting each sentence through the window of an acoustic isolation room, so that it appears on the opposite wall of the room. The talker, situated in the center of the room, was instructed to read the displayed sentence as if it were something he wished to say. This procedure avoided reflecting surfaces being near the talker, and avoided awkward diversions for the talker, such as his having to handle large decks of cards with sentences written on them.

A large set of hypotheses has been compiled, specifying how prosodic features relate to various linguistic structures. One set of hypotheses relates to how boundaries between certain syntactic units are evidenced by fall-rise patterns in fundamental frequency contours. It is hypothesized that the rise in fundamental frequency will begin at the first stressed syllable of the following syntactic constituent. The first subset of data base sentences to be processed and analyzed deals with this question of how boundaries move as the first stress in a constituent moves from the first to subsequent syllables. This subset involves 99 sentence structures of fairly similar form, but with minimal contrasts from sentence to sentence. As of the time of writing this report, all 99 sentences spoken by one talker, and 37 of them spoken by each of two other talkers, were digitized and processed through a fundamental frequency tracker and a program for detecting boundaries at substantial fall-rise patterns in the fundamental frequency contour. Processing errors resulted in 14 sentences being unavailable, so that results for a total of 159 sentences were available.

The results from this first subset indicate that, while there is not conclusive evidence as to which syntactic constituents are separated by fall-rise patterns of fundamental frequency, it is true that the rise will begin at the first stress in the following constituent. A boundary between constituents will thus appear to move more and more into the following constituent as the first stress moves to later points in the constituent. While boundaries will occasionally appear within noun phrases (such as between adjectives and nouns), it is not yet clear what causes those boundaries to appear sometimes but not always for the same structures.

An interesting sidelight to our studies of fundamental frequency contours in the 159 sentences was the observation that large fundamental frequency variations occurred before many stressed word-initial vowels. These were obviously the result of glottal stops. The glottal stop is often preceded by a local rise in fundamental frequency,

which suggests an acoustic (hence, universal physiological) origin of rising tones that are often found to precede glottal stops in tone languages. After a glottal stop, a rapidly rising fundamental frequency, or other major perturbations of fundamental frequency, may occur. Unvoicing may be apparent during the glottal stop. Obviously, fundamental frequency variations thus may be indicative of the occurrences of glottal stops, so that they may be distinguishable from oral stops.

In addition, the results with the 159 sentences strongly indicate that glottal stops are more likely to occur before stressed vowels than unstressed ones. If a glottal stop occurs, it very probably precedes a stressed vowel, and is often likely to be just after a major constituent boundary. The glottal stop is thus another potential cue to stress and constituent structure.

We thus have several potentially useful cues to constituent boundaries: fall-rise patterns in fundamental frequency contours; lengthened vowels and consonants in phrase-final positions; long time intervals between stresses; pauses; and occurrences of glottal stops.

These initial studies will soon be extended to all 99 sentences spoken by all three talkers, then to other sentences which test stress movements and boundary positions within other sentence structures. Then, subsequent studies will be conducted on subsets of sentences dealing with: (1) prosodic patterns associated with various types of sentences (yes/no questions, commands, declaratives, or "WH-questions"); and (2) subordination and bracketing.

Sperry Univac will be supplying refinements of current computer programs for prosodic analysis, plus development of new programs (such as for determining rhythm and rate of speech). We are just beginning an investigation of how to use syntactic boundaries and other prosodic information to aid the parser in the Bolt Beranek and Newman speech understanding system. Similar work on other prosodic aids to speech understanding systems is planned.

TABLE OF CONTENTS

1. INTRODUCTION

This is a Final Report on Sperry Univac's third contract under the Speech Understanding Research program sponsored by the Advanced Research Project Agency (ARPA). Sperry Univac's research is concerned with extracting prosodic information from the acoustic waveform of connected speech (sentences and discourses), and using that prosodic information to detect phrases boundaries, locate stressed syllables, determine rhythm and rate of speech, and apply such prosodic features to guiding word matching, syntactic parsing, and semantic analyses.

Under two previous contracts, Sperry Univac has performed a series of experiments on prosodic phenomena, developed some tools for extracting prosodic information from the speech signal, and played an active role in contributing to the cooperative efforts of all contractors under the ARPA SUR Program. Table I lists these previous contributions. Our experiments and algorithm developments have been published in speech journals and described at speech conferences, and have formed a basis for considerable research by other groups (Cheung, 1974; Cheung, et al, 1975; Cheung and Minifie, 1975; Maeda, 1974; Minifie, 1975; Minifie and Cheung, 1975; Sargent, 1974). Also, our prosodic analysis tools continue to contribute to speech understanding systems. Toby Skinner's autocorrelation method for fundamental frequency tracking has been implemented at several other research facilities, and versions are currently operational within the BBN and SDC speech understanding systems (Woods, et al, 1975a; Gillman, 1975), as well as within Sperry Univac's speech recognition and word spotting systems (Kloker, 1975; Skinner, 1975). Lea's programs for detecting boundaries between syntactic constituents and locating stressed syllables (Lea and Kloker, 1975), which will be summarized in Section 2 of this report, have been delivered to ARPA contractors, and are currently operational (and being tested) in the BBN speech understanding system (Woods, et al, 1975a, b).

It should thus be apparent that Sperry Univac's research has been basically two-pronged: (1) developing prosodic analysis tools and providing other services to speech understanding systems builders; and (2) conducting experiments suitable for determining exactly how prosodic patterns relate to sentence structures. The experiments have a definite practical goal in mind, however. They are intended to provide adequate understanding so that prosodic tools can be implemented and improved in such a way as to provide substantial aids to other aspects of the speech understanding process.

TABLE I. CONTRIBUTIONS OF SPERRY UNIVAC TO THE ARPA SPEECH UNDERSTANDING RESEARCH PROGRAM (1972-1974)

EXPERIMENTAL RESULTS

- Listeners can reliably perceive which are the stressed syllables in connected speech, with only 5% variations from time to time or listener to listener. This provides a reliable standard for evaluating procedures for automatically locating stressed syllables.
- 90% of all major syntactic boundaries were accompanied by fall-rise "volleys" in fundamental frequency contours.
- 85% of all syllables which listeners perceive as stressed were located by hand analysis with an "archetype contour algorithm", which used a combination of prosodic cues (increases in fundamental frequency, durations of syllabic nuclei, and energy levels).
- Simple computer programs for locating syllables from energy contours alone or fundamental frequency contours alone failed to locate many of the perceived stresses that the "archetype contour algorithm" located, and introduced more false locations than that algorithm.
- Various methods for automatic phonetic segmentation and labelling of speech were shown to produce far fewer errors in vowel and obstruent classification in the stressed syllables than in unstressed or reduced syllables. Other reasons were demonstrated for giving special attention to stressed syllables in speech understanding systems.
- Stressed syllables tend to be spaced about 0.5 seconds apart, but the time interval is a direct function of the number of intervening unstressed syllables. Pauses at clause and sentence boundaries are one- and two-unit interruptions of the rhythmic occurrence of stressed vowel onsets, respectively.
- A data base of 1100 sentences was designed to carefully isolate factors influencing prosodic and phonetic structures. A set of 178 "Phonetic Sentences" are especially suitable for testing automatic schemes for formant tracking, phonetic segment classification, and phonological rules application. A set of 922 "Prososyntactic Sentences" was designed such that various minimal pairs of sentences could isolate prosodic effects due to sentence type, syntactic bracketing, subordination, coordination, lexical stress patterns, semantic contrasts and phonetic sequences. The data base includes sentences typical of those to be handled by the ARPA speech understanding systems.

AIDS TO SPEECH UNDERSTANDING SYSTEMS (SUS's)

- Our syntactic analysis of 250 sentences produced by SUS contractors, resulted in the selection of 27 sentences which formed the large part of a data base of 31 "ARPA Sentences" used in various common studies such as workshops on parameterization, speech segmentation, and phonological rules.
- A program for fundamental frequency tracking was developed, supplied to ARPA contractors, and implemented in versions on the BBN and SDC SUS's.
- Sperry Univac and the ARPA contractors have cooperated in major common tasks of selecting speech data bases, standardizing recording procedures and phonemic notations, compiling and applying sound structure rules, comparing speech parameterization techniques and speech segmentation results, and other comparative activities.
- Sentences which had been troublesome to the BBN speech understanding system were processed through the Sperry Univac prosodic analysis programs, and specific prosodic cues were found that could be used to determine the type of sentence and the specific syntactic bracketing intended by the talker.
- An overall strategy for prosodically-guided speech understanding has been specified.

In Section 2 of this report, the latest in our series of prosodic analysis tools are briefly summarized. These programs (one for detecting syntactic boundaries and the other for locating stressed syllables) were described more fully in our recent semi-annual technical report (Lea and Kloker, 1975). Some minor improvements have been implemented since that earlier report, and the programs have been integrated into the total acoustic analysis module of the Sperry Univac speech research facility.

In Section 3 of this report, our first experiments with an extensive speech data base are described. It is important to realize that Sperry Univac's previous experiments (including those to be described in Section 3.1) have been basically "natural experiments" (Anderson, 1966), in which we did not directly control an independent variable (such as syntactic bracketing) and study resultant changes in a dependent variable (such as valleys in F_0 contours). Instead, we simply looked at the data obtained from naturally-occurring phenomena (such as the speech previously recorded and identified as the Rainbow Script, Monosyllabic Script, and the 31 ARPA Sentences). Now we are beginning a series of <u>controlled experiments</u>, in which all (or almost all) variables except one are fixed in the comparison of two utterances. These experiments provide the proper extension from the encouraging results of the natural experiments, to permit determining some well-defined rules relating prosodic variables and linguistic structure. In particular, the first subset of the designed data base has been processed, involving 173 sentences which test how F_0 detected phrase boundaries move when the first stress in a constituent moves.

In Section 4, we present the major conclusions from our experiments and our work on prosodic components for speech understanding systems. We also outline further experiments to be done with the data base, further refinements and extensions to the prosodic analysis tools, and plans for directly aiding speech understanding systems by incorporating prosodic information in parsing and word matching procedures.

References are given in Section 5, and in Section 6 an Appendix lists the sentences in the first subset of the data base, as they were processed for this report.

2. COMPUTER PROGRAMS FOR PROSODIC ANALYSIS

2.1 Prosodic Programs Delivered to ARPA Systems Contractors

Two computer programs for prosodic analysis have been delivered to ARPA contractors. One program ("BOUND3") is an improved procedure for detecting boundaries between major syntactic phrases from substantial fall-rise "valleys" in the contours of fundamental frequency versus time (Lea and Kloker, 1975). This program has been improved from earlier Sperry Univac versions (Lea, 1973a), by using more efficient procedures for finding valleys in the fundamental frequency contour, eliminating some false boundary detections by more strict requirements on the durations of falls or rises in fundamental frequency, and assigning confidence measures to each boundary detection.

The other program ("STRESS") represents a major milestone in Sperry Univac's efforts to provide prosodic aids to speech understanding. It is an implementation of a procedure for locating stressed syllables in continuous speech. This program includes procedures for finding the high-energy nucleus of each syllable in the speech and measuring the 'size' (energy and duration) of each nucleus. Those syllabic nuclei that are stressed are then found by a context-dependent analysis of energy and fundamental frequency (F_O) contours, within each major syntactic constituent delimited by the boundary detection program. Stresses are assumed to be associated with high energy (long duration) nuclei near either the peak fundamental frequency in each constituent or subsequent regions where fundamental frequency rises above a gradually falling "archetype F_O contour."

The STRESS program implements the procedures used in our previous hand analyses of stress patterns (Lea, 1973a). However, a number of improvements and new tests are included in this computer implementation of the "archetype-contour algorithm." For one thing, two different measures of the size of syllabic nuclei are used, and allowance is made for cases when extreme values of energy alone or substantially rising fundamental frequency alone may cause a nucleus to be chosen as the stressed syllable. Also, when the archetype line covers a long time span and fundamental frequency drops substantially below the archetype line, an additional test allows stresses to be found on long-duration nuclei, even if the fundamental frequency doesn't rise above the archetype line. An additional test permits long-duration nuclei just before pauses in the speech to be found as stressed.

The STRESS program was tested with several speech texts for which we already had listener's perceptions of stress levels, plus results of applying two simpler stress location programs and the hand analysis with the archetype algorithm (cf. Lea and Kloker, 1975). On the average, 89% of the syllables perceived as stressed were found by the program, while about one out of five locations were 'false', in that they did not locate a syllable perceived as stressed. Over half of these false locations were found to be due to fairly prominent ('almost stressed') syllables, false boundary detections, and failures in syllabic segmentation. However, other errors were due to detailed inadequacies in the STRESS program, such as the wrong choice of candidate nuclei, problems with the archetype line, some long prepausal unstressed syllables which appeared stressed, and short nuclei and falling fundamental frequency contours that resulted from unvoiced obstruents surrounding short stressed vowels.

It is important to note that, while the STRESS program confuses about 15% of all syllables between the "stressed" and "unstressed" categories, listeners at their best performance confuse 5% of the syllables. Thus, while the program is open to some improvements, it is approaching the level of performance that listeners can attain. The program has also been shown to work considerably better than some simpler stress location programs (Lea and Kloker, 1975).

2.2 Integration of Prosodic Programs into the Sperry Univac System

When BOUND3 and STRESS were first implemented and supplied to ARPA contractors, they were independently operating programs. BOUND3 operated on F_0 contours obtained from cards or card-image files. Although STRESS used the boundary positions and other information determined by BOUND3, it read those values from data cards, rather than from stored arrays. STRESS printed out a lot of results, including diagnostic statements. Small input and output changes were needed in both BOUND3 and STRESS, to get data and results in the forms desired in specific systems and to have the programs operate together, within the total structure of a speech analysis system.

Recently, the BOUND3 and STRESS programs have been integrated into the acoustic analysis module of the Sperry Univac Speech Research Facility. As shown in Figure 1, this module includes programs for F_0 tracking, extraction of sonorant energy from a band limited (60 Hz to 3000 Hz) integration of the energy in the LPC spectrum, and a voicing decision (based on the energy in the frequency band 60 to 400 Hz). BOUND3 uses only the F_0 contour (in the form of an accessed data file), to obtain positions of

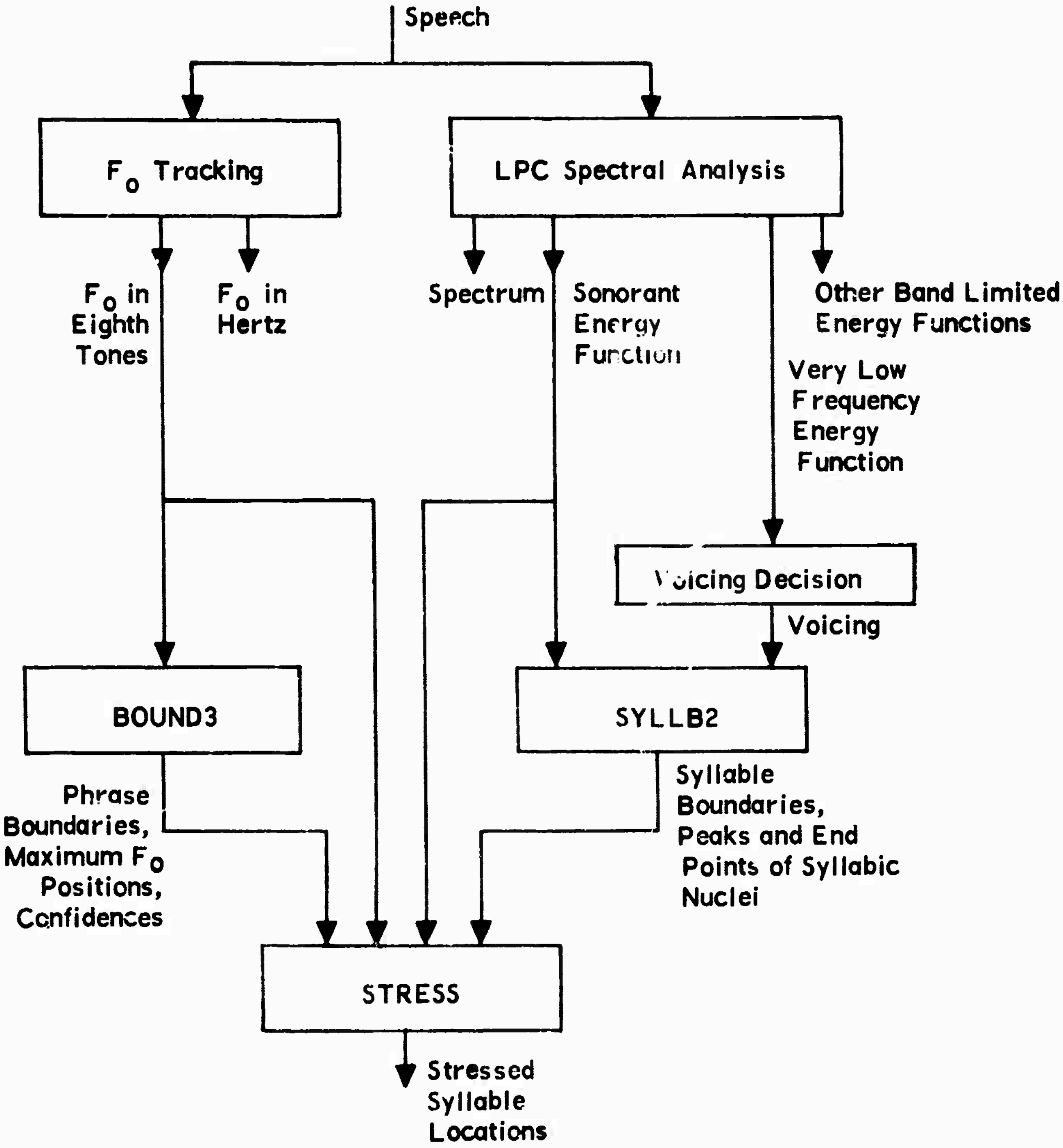


Figure 1. The Prosodic Analysis Programs, as Integrated into the Sperry Univac Acoustic Analysis Module

constituent boundaries, maximum-F_0 points in each constituent, and confidences. As shown in Figure 2, the SYLLB2 program uses the regions of high sonorant energy that are voiced, to locate the beginning, end, and highest-energy points for each syllabic nucleus, plus the energy dips between syllables, which constitute syllable boundaries. Then, the data files of constituent boundary information and syllable information are used by STRESS to locate stressed syllables.

One significant refinement in this system structure is that syllabification is now an independent process which precedes stress analysis, rather than a part of the process of stressed syllable location, such as it was when the STRESS program was first developed. The syllable information can be used by other analysis procedures, such as in the phonetic analysis module which locates vowels and consonants in the utterance. Lexical matching procedures in a speech recognition system can also use the syllable boundaries (cf. Kloker, 1975).

2.3 Cooperative Work with ARPA Contractors

Listings and computer-readable forms (cards and digital tapes) of the BOUND3 and STRESS programs were provided to ARPA systems contractors, along with some sample results for a few speech data files. BBN made both programs available over the ARPANET. BBN also integrated both programs with their version of the Sperry Univac F_0 tracking algorithm and the BBN procedures for finding sonorant energy and voicing. BBN and Sperry Univac interacted on the evaluation of the boundary detection program's performance with ten BBN sentences. This work is described more fully in a BBN quarterly progress report (Woods, et al., 1975a).

In addition to providing computer programs for prosodic analysis, Sperry Univac has participated this year in various cooperative activities with other ARPA contractors. Wayne Lea presented a talk on Prosodic Rules and Hypotheses at the ARPA Phonological Rules Workshop held at BBN. Mark Medress has participated in regular activities of the ARPA SUR Steering Committee, and has been appointed the Associate Chairman of the steering committee. Sperry Univac also actively participated in the drafting of a Follow-On plan for further speech understanding studies following the current five year program. Most recently, Sperry Univac has begun cooperative studies with BBN to determine how prosodic information might be used to aid the BBN parsing procedures. This will be discussed further in Section 4.4.

3. EXPERIMENTS ON PROSODIC CUES TO LINGUISTIC STRUCTURE

3.1 Timing Cues to Linguistic Structure

3.1.1 Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries

During this contract year, a series of experiments have been conducted on timing cues to linguistic structure (Lea and Kloker, 1975). In one experiment (Kloker, 1975), five sentences per speaker were selected from the speech of six individuals who participated in simulations of computer interactions. The utterances were distorted by spectral inversion and presented to five listeners who marked stressed syllables and the locations and types (normal or hesitation) of phonological phrase boundaries, using only the prosodic cues remaining in the signal. Vowel and sonorant durations (with and without aspiration) were measured from spectrograms, and then declared stressed or unstressed based on the perceptions. Exploring the hypothesis that large increases in phonetic duration are syntactically determined, perceived boundary locations were compared with preceding segments which were 20% above the median length for that segment type. Using a rule which groups lengthened syllables, and from the lengthened group predicts phrase boundaries, 91% of the perceived boundaries were predicted. Of all the perceived phrase boundaries, those before silences longer than 200 milliseconds were more reliably predicted by lengthening than boundaries not at long silences. Locations perceived to be normal phonological phrase boundaries were more reliably predicted than those perceived as hesitations. Of the predicted boundary locations not perceived by listeners, some mark major syntactic boundaries, but most are at minor syntactic breaks, notably between modifiers and nouns, and after prepositions. The results also suggest that speaker differences and style variations may be important.

3.1.2 Interstress Intervals as Cues to Phonological Phrase Boundaries

In another experiment (Lea, 1975), the question was whether or not one could detect major phrase boundaries from timing of prosodic features alone (such as onsets of syllabic nuclei found from energy contours), without the need for a prior determination of the pnonetic sequence or the detection of lengthening of phonetic segments. We found that, in read sentences and paragraphs, as well as simulated man-computer interactions, time intervals between onsets of stressed vowels ("disjunctures") clustered near mean values around 0.4 to 0.5 second, with standard deviations of about 0.2 second. Contrary to published hypotheses, durations of disjunctures tended to increase about linearly with

the number of intervening unstressed syllables. Mean disjuncture durations doubled when spanning clause boundaries, and tripled when spanning sentence boundaries. Mean pause durations, as measured by durations of unvoicing, tended to be equal to or twice the mean interstress interval, for clause and sentence boundaries, respectively. Syntactically-dictated pauses thus appear to be one- or two-unit interruptions of rhythm. Long disjunctures also accompanied 95% of the perceived boundaries between phonological phrases, and were found useful in determining which of several minimally-contrastive syntactic structures had been spoken

3 1.3 Interstress Intervals as Cues to Applicable Phonological Rules

In a third experiment (Lea, 1975), we investigated how various measures of the rate of speech correspond with changes in phonological structure that should be handled by "fast speech" phonological and acoustic phonetic rules. The duration of the interstress interval was found to inversely correlate with the percentage of phones that were erroneously categorized by various available methods for automatic phonetic categorization. Other measures of speech rate, such as the number of syllables per unit time, were not as closely correlated with phonetic error rates. The interstress interval thus appears to be useful in predicting phonological rules that might apply to an utterance.

These experiments show further ways in which the location of stressed syllables could play an important role in speech understanding, and they expand the ways in which prosodic information could be used to determine syntactic and phonological structure.

3.2 Design and Recording of Speech Data Base

3.2.1 Data Base Design

There is a definite need to develop precise rules for systematically relating prosodic patterns to underlying linguistic structures. Studies of "near-minimal-pairs" of sentences, which are identical except for one difference in linguistic structure (or a few isolatable differences in structure). will help determine the correct form of the rules relating prosodic patterns to linguistic structures. It is for such reasons that a data base of 1100 sentences was designed near the end of our previous contract (Lea, 1974a,b). It includes 922 "Prososyntactic Sentences" which are designed to explicitly test the prosodic effects of sentence type, contrastive syntactic bracketing, subordination, coordination, syntactic categories (such as pronouns, verbals, compound nouns, etc.), movement of stress within phrases, coreference, etc. Prosodic patterns to be studied for these

sentences include: performance of the program for detecting phrase boundaries from valleys in F_0 contours; acoustic correlates of stressed syllables, and performance in automatic stressed syllable location; acoustic measures of rhythm and rate of speech; overall F_0 contour shapes; and local variations in prosodic features due to phonetic sequences. (In Section 3.3, 3.5, and 3.6, we will discuss some specific hypotheses that are currently being tested or that will be tested during our next contract.)

In addition to the F[illegible]syntactic Sentences, the data base includes a set of 178 sentences which include all word-initial consonant-vowel (CV) sequences, and all word-final vowel-consonant (VC) sequences. These "Phonetic Sentences" provide the speech data needed for efficiently testing automatic procedures for vowel and consonant classification. For example, five sentences provide instances of all distinguishable stressed vowels of American English, coupled with the sibilants [s, ʃ], in initial CV and final VC positions.

From extensive studies with such designed sentences, we hope to develop experimentally-validated intonation rules and other prosodic and phonological rules. These rules will then be used to guide parsing, semantic analysis, phonological analyses, and word matching procedures in ARPA speech understanding systems.

3.2.2 Ordering the Sentences

The minimal contrasts that exist between many of the designed sentences require a careful ordering of the sentences, to prevent talkers from introducing contrastive stress and other undesired comparisons between sentences spoken one after another. Also, since about three-quarters of the data base sentences are all-sonorant (that is, they contain no obstruents), while the Phonetic Sentences and some task-related sentences include obstruents, it is desirable to intersperse an obstruent sentence about every fourth sentence. This breaks up the all-sonorant pattern, hopefully preventing any extreme tendencies to articulate in some unusual manner.

To intersperse the obstruent sentences with the sonorant ones, and to eliminate sequences of very similar (obviously contrastive) sentences, the following procedure was used. First, the set of 1100 sentences was divided into four equal-size subsets (each subset thus containing 275 sentences). One subset included all the 178 Phonetic

Sentences, plus 97 other sentences (primarily task-oriented sentences) which contained obstruents. The other three subsets contained short, medium, and long all-sonorant sentences, respectively.

The sentences within each of the subsets were then numbered, with the simpler sentences first, and the more complex sentences in the subset being last (i.e., with the largest numbers). Then the 275 "short" sentences were randomly ordered, by the use of a random number table. The resulting random number list (S104, S223, S241, S094, ..., S217, S084, S123, S085) dictated that short (S) sentence 104 would appear first, then short sentence S223, etc. The random order prevented sentences that were even four apart from each other in the final list from being similar, so that S2 didn't follow S1, etc. Then, to further reduce the likelihood of very similar sentences being adjacent, the long (L) sentences were ordered in the reverse of the order for the short sentences (L085, L123, L084, L217, ..., L094, L241, L223, L104). The medium sentences were ordered beginning at the 101st number in the order of the short sentences, progressing to the end, then concluding with the first 100 of the short-sentence list. The obstruent sentences were ordered beginning at the 101st number in the order of the long sentences.

By first choosing a sentence from the subset of short sentences, then one from the subset of long sentences, then one from the medium-length sentences, and then one from the subset of obstruent sentences, and repeating that pattern of short, long, medium, and obstruent sentences, we prevented sequences such as two short minimally-contrastive sentences, and also avoided any appearance of obvious patterning, such as would be obtained with progressively more complex sentences, such as: short, then medium, then long, then obstruent.

The result was a list of sentences which had very few instances of similar sentences being adjacent. There were a few cases where there was close semantic similarity between successive sentences (e.g., L224: "Either Lou or Neal will know Ron." and M104: "Will I know?"), that could introduce undesired sentence sequence effects (such as contrastive stress on "I" in M104). These were eliminated by interchanging one of the undesirably sequenced sentences with another sentence in the list, such that the resulting order had no such undesired sequences. A trial run of reading all the sentences out loud convinced the experimenter that no undesired sequences remained.

This careful effort in ordering the sentences prevents undesirably introducing discourse effects such as contrastive stress into single sentences, when the relationship

between the syntax and the prosodies of each sentence is being tested. (Also, obvious phonetic contrasts, such as a sentence with lots of / s /'s following one with many / z /'s, are avoided by the ordering.) If a researcher should be interested in directly studying discourse effects such as contrastive stress, it is easy to develop lists of sentence pairs that would be suitable. Also, paragraphs can be composed from the sentences, to study supra-sentential effects such as paragraph intonation, prosodic cues to coreference, etc. For our immediate studies, these supra-sentential effects are purposely avoided, so that clearer associations between the syntax and prosodies within individual sentences could be determined.

3.2.3 Equipment and Talkers

High quality recordings of the total set of 1100 sentences were obtained with a Shuir Model 548 microphone with windscreen, Scotch 208 analog tape, and a Sony TC-650 tape recorder, for three male talkers individually situated in an IAC Moduline (12 foot by 12 foot) acoustic enclosure. The recording heads on the tape recorder are cleaned each day (after each 4 to 6 hours use). To date, only one reading of all 1100 sentences by each of the three male talkers has been recorded, although our original plans (Lea, 1974c) called for the later expansion to three repetitions by each of the eight talkers, with repetitions spaced in time by at least one week.

The three male talkers were selected from ten available talkers whose acoustic data (fundamental frequency contours, sonorant energy contours, formant tracks, etc.) had been determined for other connected speech utterances. The selected talkers had exhibited clear, consistent articulation, clear formant tracks, apparently normal prosodic pronunciation, and a variety of pitch levels.

All three talkers are of similar mid-Western American background, with no prominent dialectal pecularities. A dialectal background questionaire was used to establish regional, social, and educational factors that might affect pronunciation. The author is indebted to Dr. Paul S. Cohen for providing the questionaire. The questionaire was administered to each talker orally, and the questions and responses were tape recorded, for possible studies of how the talkers speak with interview and casual speech, as well as with read speech.

It is significant that one of the three talkers was the experimenter, who is also the author of the sentences. While his knowledge of the purpose of the data base and

his background in studying speech, linguistics, and prosodic structures may add a substantial controlled bias to his method of pronunciation, it did se reasonable to have one talker aware of what prosodic, syntactic, and phonetic contrasts are being studied, and what constitutes good pronunciation. The other two talkers were not familiar with the detailed purposes of the sentences, and have not been trained in speech or linguistics.

3.2.4 Presentation of Sentences to the Talker

In previous recordings of small lists of isolated words, phrases, or sentences, we have either had the talker read individual utterances from cards which he handled, or we had the experimenter display a card at the window of the acoustic enclosure, and the talker read utterances as they are thus presented. For this large data base, we concluded that more sophisticated procedures had to be used, so the talker did not have to handle large decks of cards, and the reflective surface of the window was not directly in front of the talker.

The chosen procedure for recording this data base had the experimenter sitting outside the window of the quiet room, operating the tape recorder and determining the order and timing of sentence presentation. The talker was situated in the middle of the quiet room, facing the wall opposite the window. (The talker is then about six feet from the wall and other reflective surfaces.) A single sentence is projected through the window onto the opposite wall, using a 35mm slide projector and 1100 slides (clear lettering in a dark background). The talker was instructed to read the sentence through quietly first, to be sure of what he was to say, and then to speak the sentence naturally, without special emphasis on any word, and without contrasting that sentence with any other he might have previously spoken. If the talker felt he made a mistake in saying a sentence, he was permitted to repeat it. If the experimenter concluded that there was any problem with the pronounced sentence, he marked it on his list, for recording again at the end of the day's recording for that talker.

At the end of each tape (about every 30 minutes), the talker was given a five minute break to relax and stretch, and to permit a change of recording tape. Each tape contained 160 sentences (from two carousels of 35mm slides) with about 5 or 6 sentences spoken per minute. Recording sessions were limited to about two hours per day per talker. Two or three talkers could be recorded on about half the data base per day. The remainder of the sentences for each talker were then recorded on another day.

3.3 Compiling a List of Prosodic Hypotheses

We now have a carefully designed and recorded data base that should help us better understand how the interacting effects of semantics, syntax, lexical structures, stress patterns, and phonetic sequences are superimposed in the F_0 and energy contours of controlled English sentences. Coupling such data with some explicit hypotheses about what regularities to look for in the data, we then should have some of the most essential understanding needed to use acoustic prosodic data in guiding speech understanding strategies. Our previous natural experiments have suggested some useful prosodic regularities, and many published (but untested) hypotheses are available, yet more precise rules or hypotheses are clearly needed. For example, our previous predictions of where phrase boundaries should occur in F_0 contours have been based on intuitive analyses of syntactic structures. Where expected boundaries do not occur, or wherever false or unexpected boundaries do occur, there has been no recourse indicating the source of the error. This is in part due to the intuitive predictions used, and in part due to the uncontrolled syntactic structures involved in the texts studied. Experiments with the designed sentences of known syntactic structure (see Section 3.4 to 3.6) will indicate exactly what structural boundaries are marked, and will permit the writing of precise rules predicting where boundaries will occur in new sentences of similar structures. These rules for predicting detectable boundaries may then be useful in computer determination of possible underlying structure, given the detected boundaries.

Similarly, precise rules for relating stress patterns to underlying structures are needed. Published (theoretical) hypotheses about sentence stress patterns need to be tested with experimental data, so that we ultimately can write reliable analytical procedures for predicting underlying structures from stress patterns.

A careful study of the literature, and previous analyses of prosodic data, have resulted in the compilation of an extensive set of hypotheses and rules relating prosodic patterns to linguistic structures. These will be summarized in a forthcoming report (Lea, to appear). Here we will consider only a few hypotheses related to the occurrences of valleys in the F_0 contour near major syntactic boundaries. These hypotheses will be tested with the first subset of data base sentences.

We begin with a hypothesis that explains our use of all-sonorant sentences in the first subset:

Hypothesis A: There are no "substantial"[1] variations in F_0 values due to occurrences of various categories of sonorant consonants and vowels.

That is, while obstruents may cause substantial (10% or more) jumps and dips in F_0 values (Lea, 1973b), sonorant sounds are expected to produce smooth monotonic contours. Consequently, valleys in F_0 contours of all-sonorant sentences will not be due to phonetic effects, but rather due to stress and syntax patterns, as some later hypotheses will specify.

The major hypothesis about the occurrence of F_0 valleys marking syntactic boundaries is the following:

Hypothesis B: Each boundary between two "major grammatical constituents" will be marked by a "substantial"[1] decrease in F_0, followed by a substantial increase in F_0 (that is, a substantial F_0 "valley").

Some of the terms and conditions in this hypothesis need further definition, as indicated by the following hypotheses:

Hypothesis C: "Major grammatical constituents" (in hypothesis B) include noun phrases (NP), prepositional phrases (PP), adverbial phrases (AdvP), main verbs (V), embedded clauses (S_i), and matrix sentences (S).

Hypothesis D: Boundaries will also be marked by F_0-valleys between a sentence adverb and its surrounding constituents, except that a preverbal adverb is not so separated from its following verb (as in "really enrolled").

Hypothesis E: Boundaries will also be marked by F_0-valleys between the words of a lexical compound (noun-noun like "constituent boundary" or compound-verb like "backbite", etc.).

The following hypothesis is the primary one to be tested with the first subset of the data base:

Hypothesis F: The bottom of an F_0 valley (which in Lea's previous work has been declared as the "position" of the detected syntactic boundary) will occur just before the first stressed syllable in the latter constituent.

[1] "Substantial" variations in F_0, while they may be text and talker-dependent, are expected to be on the order of 5 to 10% or more, based on Lea's earlier studies (Lea, 1972; 1973a,c).

Hypothesis F asserts that the position of a boundary will be just before the first stress in a constituent, so that boundary detections might ultimately be related to boundary locations by an understanding of this systematic displacement of detected boundaries. (One exception to this positioning of boundaries should be noted: when the earlier of two major constituents ends in a "Tune II" intonational rise, as a mark of incompletion, then the boundary may occur just before the end of the earlier constituent, within the valley produced just before the phrase-final increase in F_0.)

Some useful corollaries follow from hypothesis F (and the set of related hypotheses B to E):

Corollary F1: A boundary will move more and more "into" the following constituent as the first stress moves to later points in the constituent.

Corollary F2: A boundary will be marked (detected by an F_0 - valley) before (or within the beginning of) a constituent only if that constituent contains a stressed syllable.

Corollary F3: The F_0 valley will appear before the first stress in a constituent, regardless of whether or not that stress appears in the first word of the constituent.

One syntactic constituent which hypothesis C does not mention is the auxiliary verbal constituent, which has been the topic of considerable recent research (Allen, 1973, Allen and O'Shaughnessy, 1974). To predict the F_0 contour effects associated with auxiliary verbs, we need a few hypotheses about stress patterns in various words and constituents.

Hypothesis G: "Auxiliary" or "function" words are normally unstressed. These include pronouns, articles, prepositions, conjunctions, "that" – complements, auxiliary verbs, and the word "there" resulting from the "there – insertion" transformation.

Hypothesis H: Lexical words (nouns, verbs, adverbs, adjectives) have one stress. Secondary or tertiary stresses in a word are not normally heard as stressed, nor do they show acoustic correlates suggesting stress.

From hypothesis G and corollary F2, we derive the following:

Corollary G1: A constituent boundary will not be marked between a noun phrase subject of a sentence and its following auxiliary, unless that auxiliary contains a stressed syllable (as when emphasis is added, or negation is introduced into the auxiliary).

This corollary, if true, would explain why boundaries have usually not been detected between noun phrase subjects and following auxiliary verbs (Lea, 1972, 1973a, 1973c).

The word "normally" is used in the hypotheses G and H to refer to non-emphatic speech with no special semantic connotations such as emphasis, emotion, etc. Agreeing with Chomsky and Halle (1968), Kurath (1964), Trager and Smith (1951), and many others, and rejecting some weak claims by Bolinger (1965), I contend that there is something which we can call normal, unmarked, syntactically – dictated prosodic patterns, and these are to be distinguished from deviant cases where syntax and prosodies conflict (as when declarative word sequences are combined with the rising intonation of a yes-no question) or where special semantic moods and speaker attitudes are reflected in prosodic structure. Perhaps this should be considered as a global hypothesis relevant to the designed sentences:

Hypothesis I: There are "normal", "unmarked" prosodic patterns which are dictated by lexical and syntactic structure, and which may be evident in readings of the designed sentences.

If some individual talkers actually convey semantically-marked versions of the sentences, such as emphasizing "one" in a sentence "$\overset{S}{\text{Ron}}\ \overset{S}{\text{knew}}\ \overset{EMPS}{\text{one.}}$", when a normal expected pattern may be "$\overset{S}{\text{Ron}}\ \overset{S}{\text{knew}}\ \overset{U}{\text{one.}}$", that does not refute the existence of "normal" patterns.

For each of the prosodic hypotheses, one can devise counter-hypotheses that refute them. In particular, one counter-hypothesis that conflicts with hypotheses B, C, and F is the following one:

Hypothesis J: An F_0 valley will occur just before every stressed syllable, regardless of whether it is the first stress in a constituent or not.

Past results suggest that hypothesis J is wrong. To test it, we need consider cases of constituents that contain more than one stress. One category of constituent that can contain more than one stress is the noun phrase, which is expanded to a variety of forms in the first subset of the Sperry Univac data base. Bolinger (1965, pp. 57-69) has published several contentions about the F_0 contours (or, "pitch accents") accompanying various noun phrase constructions, including the following:

Hypothesis K: Quantifier – plus – noun combinations are accompanied by rising of F_0 into the quantifier and noun ("Accent B"), and falling F_0 after the noun ("Accent A"). Thus, the combination shows a single rise-fall unit pattern ("Accent B" followed by "Accent A").

Bolinger claims that combinations of descriptive-adjective-plus-noun will not show the same intonation as the quantifier-plus-noun, as follows:

Hypothesis L: A descriptive-adjective-plus-noun will be accompanied by an F_0 valley between the two words, and an F_0 valley after the noun. ("Accent A" followed by "Accent A").

I am dubious about some of Bolinger's claims, reflected in these and other hypotheses, but they will be tested with the first subsets of data base sentences. Bolinger's hypotheses may be translated into the following corollaries:

Corollary K1: A constituent boundary will not be marked between a quantifier and a noun.

Corollary L1: A constituent boundary will be marked between a descriptive adjective and a noun.

Past results suggest that corollary L1 is not quite correct. Lea (1972, p. 81) found that only about half of the adjective-noun sequences were accompanied by F_0-detected boundaries. We shall consider further results in Section 3.5, and, extend the tests to include sequences of adjective-plus-adjective, possessive-plus-noun, etc., plus more elaborate combinations like quantifier-number-adjective-adjective-noun (e.g., "any nine mean young men").

3.4 Processing a Subset of Sentences Related to Locating Constituent Boundaries

The first subset of data base sentences to be analyzed are intended to help test the hypotheses listed in Section 3.3. These sentences (listed in Appendix A) particularly deal with the prosodic effects of moving the first stress in a constituent from the first to subsequent syllables. All of the sentences in the chosen subset are simple (unembedded) declarative sentences, of one of the following structures:

NP V
NP AUX V
NP V NP
or NP AUX V NP,

except for two sentences, which are of the form

NP V NP ADV
and NP V NP NP .

The first constituent in all these sentences is monosyllabic and stressed, since we are only interested in stress movement within the other, non-initial constituents, which can be preceded by F_0-detected boundaries. By contrasting sentences with verbs like "know", "worry", and "enroll", we can determine how boundaries move as stress moves within verbs. Similarly, contrasting the F_0 contours in object noun phrases like "Ron", "Mary", "Marie", "Maria", and "Leonora", we can determine how F_0 valleys move as stress moves within the noun phrase. Single-word noun phrases like "Maria" may be contrasted with multiple-word constituents of the same stress pattern, like "an airman".

The first 24 sentences shown in Appendix A have only one stress per constituent. The next 13 have two or three stresses (in various positions) in the noun phrase, but with the noun phrase structure confined to a form with a determiner (number, possessive, quantifier, or some combination of such) plus noun. The next 54 sentences permit descriptive adjectives, participles, and adverbs, along with determiners, to appear in the noun phrase. The last eight sentence structures involve the famous "flying planes" paradigm (cf. Chomsky, 1957), in which an ambiguous sequence like "they are flying planes" is studied. However, to permit talkers to correctly speak the contrasting instances where "are V-ing" and "V-ing N" structures are intended, the wording in the eight sentences is deliberately <u>unambiguous</u>, but with pairs of similar enough phonetic

structures that comparisons can be readily made. These sentences give a simple case for testing whether prosodic structures can select among possible alternative bracketing of a sentence.

Fifty-eight of the 99 sentences in the subset involve the structure "Ron will enroll NP". In addition to testing the various NP constructions, these sentences permit extensive testing of whether or not F_0 valleys occur between subject noun phrases and auxiliary verbs, and whether or not boundaries appear just before the stressed syllable "-roll" in the verb.

The 99 sentences, as spoken by one talker (LLL), plus the first 37 sentences, as spoken by each of two other talkers (WAL and JFS) were digitized, and also dubbed onto another analog tape, in the order given in the Appendix. Digitization involved passing the signal through a seventh order elliptic function Cauer low-pass analog filter (cutoff frequency, 4782 Hz), and sampling at 10,000 points per second, with a 12-bit analog-to-digital conversion, and no hardware pre-emphasis. A separate waveform file was made for each sentence, and stored on digital tape. These digital tapes are available for use by other researchers, as are the analog dubbings.

Each of the 173 sentences were processed through the autocorrelation F_0 tracker and BOUND3 program for detecting syntactic boundaries, and plots were made of the F_0 contours with associated boundary markers. These prosodic analysis tools reside on the Sperry Univac 1108 time sharing system. Then, transcriptions were made on the prosodic plots, by listening to the digitized waveform on our Sperry Univac 1616 Speech Research Facility, and gating to determine approximate beginning and ending points of phones in the utterances. These transcriptions were needed to synchronize the F_0 contours and boundary markers with their corresponding positions in the phonetic structure of the utterance.

Unfortunately, due to some processing errors, only 159 of the 173 sentences were available for analysis at the time of writing this report. At a later time we intend to complete the F_0 processing, and to process all 173 utterances through procedures for obtaining energy contours, finding syllables, and locating the stressed syllables. Only F_0 and boundary information for the 159 sentences has been obtained to date, but that information is sufficient to test the hypotheses given in section 3.3. In sections 3.5 and 3.6 we will discuss various conclusions from the analysis of the F_0 and boundary results.

3.5 How Boundaries Are Related to the Position of the First Stress in a Constituent

Hypothesis A suggests that quite smooth F_0 contours should be exhibited for the all-sonorant sentences chosen for the first data base subset. This was indeed found to be the case (except where glottal stops occurred, as will be discussed in Section 3.6). In general, the various vowels and sonorant consonants had little effect on F_0 contours, so that the contours appeared to smoothly follow the general rise and fall shapes that appear to mark stress patterns and syntactic units. Very slight local effects due to some sonorant consonants were evident, however, and they agreed with previous observations for other speech texts. Although exact statistics have not been compiled, it appears that laterals (/ɭ/'s) tend to yield very flat local F_0 contours, there are very slight F_0 dips due to /w/'s, and nasals have fairly flat F_0 contours, with the exception of /ŋ/, which seems to be one possible source for slight F_0 dips that frequently occurred at the end of the word "young".

In general, though, there are no substantial variations in F_0 due to occurrences of various categories of sonorant consonants and vowels. Having thus successfully removed (most of) the effects of phonetic sequences on F_0 contours, we can inquire as to how stress and syntax effects are evidenced in the F_0 contours.

Our major hypotheses (B to F) state that major syntactic constituents will be accompanied by substantial F_0 valleys, and that the bottom of the valley will occur just before the first stressed syllable in the later constituent. To test these ideas, we need to define a "substantial" valley. Since F_0 generally falls throughout an utterance (Lea, 1972), a cue to stress-and-syntax-dictated F_0 variations is most evident in the F_0 rise that accompanies a stressed syllable. It is usually preceded by a substantial F_0 fall. Figure 2 shows histograms of the amount of F_0 rise that followed such phrase or word boundary in the 159 sentences.[1] Unfortunately, two of the talkers spoke with a more monotonic intonation than that exhibited by other talkers in our previous studies, so that those talkers (LLL and JFS) don't show very large F_0 rises in any instances. Still, it is apparent that some boundaries are generally accompanied by larger F_0 rises than are other boundaries The verb-noun phrase boundary (such as that between "Men worry' and "Mary" in sentence S69) exhibits the largest mean value of F_0 [illegible]. Other major syntactic boundaries before the stressed verbs (i.e., AUX-V and NP-[illegible] boundaries)

[1] Excluded from Figure 2 are 40 cases where glottal stops occurred before word-initial vowels, causing dramatic F_0 variations that are not syntax-dictated. These cases with glottal stops will be discussed in Section 3.6.

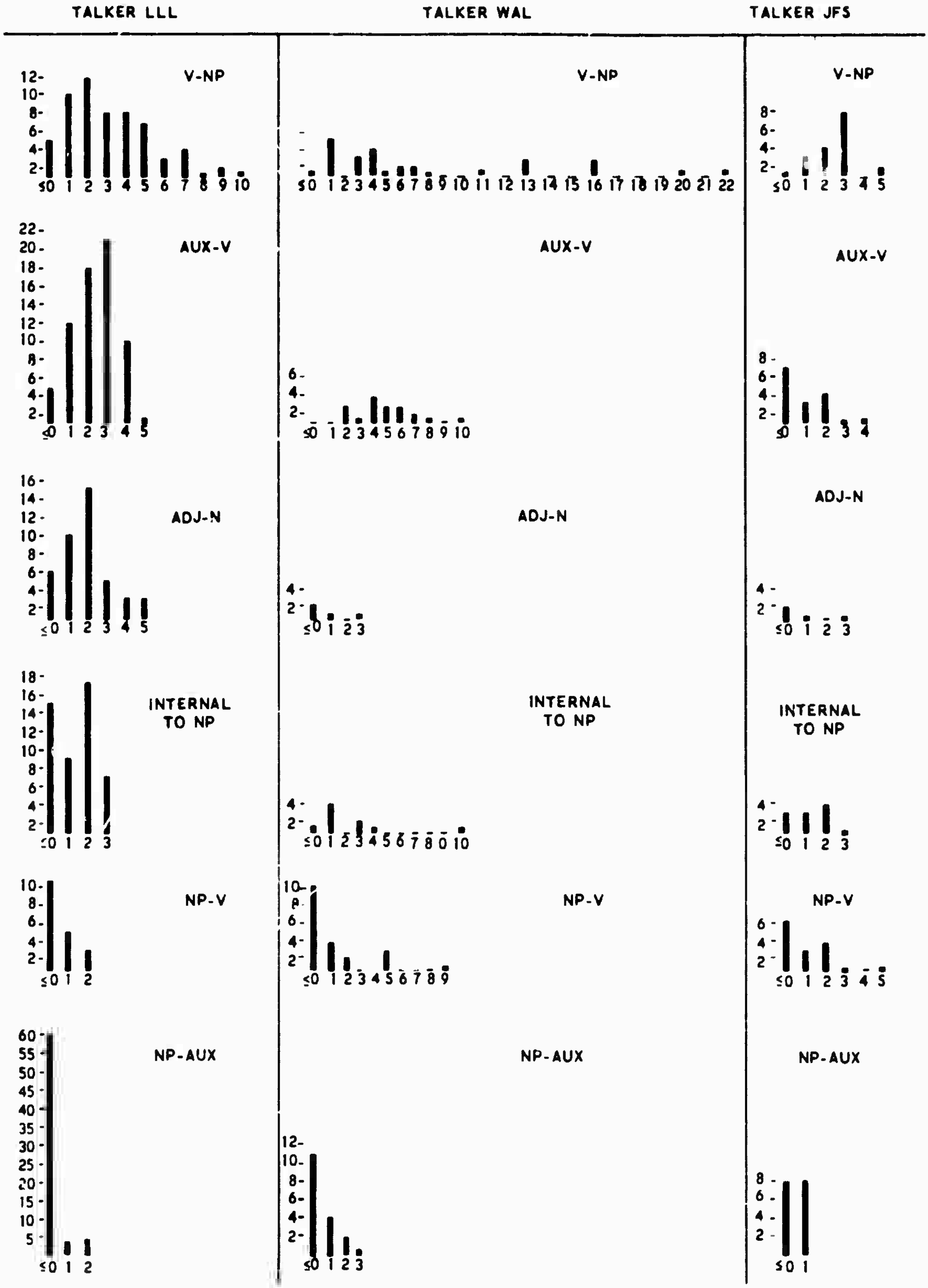


Figure 2. Numbers of Occurences (Vertical Axes) of Each Amount of Increase in F_0 (Horizontal Axes), at Various Positions in the Sentences of the First Subset of the Data Base.

exhibit F_0 rises that, on the average, are somewhat larger than those internal to the noun phrases (such as phrase-internal boundaries between quantifiers and nouns ("any men"), or between numbers and nouns ("nine men"), quantifiers and numbers ("any nine"), quantifiers and possessives ("all your"), possesives and nouns ("your men"), or quantifiers and adjectives ("all young"). Boundaries between noun phrases and (unstressed) auxiliaries exhibit little or no F_0 rises.

A simple threshold of five eighth tones rise has been used as one condition for finding "substantial" F_0 valleys in Lea's previous work with the BOUND3 program for detecting phrase boundaries. It is evident in Figure 1 that, with a very few exceptions, only major syntactic boundaries (V-NP, AUX-V, and perhaps NP-V) exhibit such a substantial F_0 rise. However, many of the major syntactic boundaries would be missed by requiring such a high threshold, whereas a lower threshold (such as only three eighth tones rise) would pick up many more of the correct boundaries, while also introducing "false" boundary detections due to F_0 variations that are not at major syntactic boundaries. It is important to recall that the large (five eighth tone) rise in F_0 was required in Lea's earlier work to help eliminate smaller F_0 variations due to voiced and unvoiced obstruents. Without such a phonetic source of erroneous F_0 variations, we would expect that a lower threshold would be possible for boundary detection within the all-sonorant sentences. For example, a simple threshold of three or more eighth tones rise required for boundary detection, for talker WAL, would correctly detect 22 of the 28 V-NP boundaries, and 17 of the 20 AUX-V boundaries. It would also "falsely" detect 5 of the 13 boundaries internal to the NP and 1 of the 18 NP-AUX boundaries. A similar threshold applied to the speech of the other talkers would yield even more confused performance (more correct boundaries missed, and more false detections that aren't at major syntactic boundaries).

Whether a "substantial" F_0 valley is identified with 5 eighth tone rises or 3 eighth tone rises, it is thus clear that hypothesis B from Section 3.3 is not invariantly true. There is a strong tendency toward major syntactic constituents being demarcated by F_0 valleys, but there are some instances where such valleys don't occur as expected, and a substantial number of instances where valleys occur which hypotheses B and C would not predict. These results are less conclusive than those found during our previous studies of F_0 valleys in uncontrolled speech (i.e., speech with obstruents, with arbitrary sentence structures, and various lengths of constituents). More studies with more sentences in the

data base, and perhaps with other talkers, may be necessary to determine exactly which constituents can reliably be expected to be accompanied by F_0 valleys. The most evident boundary is the V-NP boundary, but even with that boundary we don't yet know any systematic reasons why not all instances are accompanied by substantial F_0 rises.

These results suggest some sort of weakening or qualifications to be attached to hypotheses B and C. Further study is needed. We also don't yet have sufficient data to completely evaluate hypotheses D and E, but the few instances available suggest that NP-NP boundaries are marked, but NP-ADV ones aren't.

However, there is one hypothesis that <u>has</u> been very firmly verified with the first subset of the data base. That is hypothesis F. With very rare (and easily explainable) exceptions, all F_0 rises accompanying major phrases in the sentences <u>did</u> occur at the onsets of their first stressed syllables, not earlier or later. For example, for talker LLL, of the 110 times that any F_0 rise at all occurred near a phrase boundary (without any glottal stop), that rise occurred 107 times at the onset of the first stressed syllable in the constituent. These sentences included many instances where stress was in the first syllable of the constituent ("know", "<u>Man</u>y", "<u>air</u>men", etc.) or the second syllable ("en<u>roll</u>", "Ma<u>rie</u>", "an <u>air</u>man", "im<u>mor</u>al men"), and one case where stress was in the third syllable ("a ma<u>rine</u>", in sentence S107). One of the three exceptions was for "Leonora" in sentence S56, for which syllable "Le-" was somewhat stressed, even though "-nor-" was predicted to be the first stress in the constituent. Thus, this could well be a correct case, where F_0 rise <u>did</u> accompany the first stress, but that first stress didn't appear where expected. The other two exceptions involved small local rises in F_0 that were apparently associated with phonetic structure, not with stress or syntax. Similarly, the other two talkers exhibited F_0 rises always at the first stresses, except that both partially-stressed syllables of "Lenora" showed some F_0 rise. The case of F_0 rises accompanying the stress in "Le-" in Lenora suggests that hypothesis H may be incorrect, if one includes small F_0 rises as acoustic correlates of stress.

Thus, we conclude that, for the subset of data base sentences processed to date, <u>hypothesis F is clearly confirmed. The F_0 valley, if present, immediately precedes the first stress in the constituent.</u> As stated by corollary F1, <u>a boundary will move more and more into the following constituent as the first stress moves to later points in the constituent.</u> The F_0 valley will appear before the first stress in the constituent, <u>regardless of whether or not that stress sppears in the first word of the constituent</u>

(corollary F3). Also, the fact that NP-AUX boundaries are not accompanied by substantial F_0 rises[1] confirms corollary G1 and suggests that (as corollary F2 asserts) a boundary will be marked by an F_0 valley only if the following constituent contains a stressed syllable. We will be testing corollary F2 more later, when other unstressed constituents such as pronouns are included in the tested structures.

It was evident, both from listening to the tapes, and from the F_0 contours, that hypothesis G is being confirmed, as far as we have explored it to date. Auxiliary verbs and articles are unstressed. Also, the possessive pronoun "your" generally appears to be unstressed, even though our original stress assignments in designing the sentences predicted that it would be stressed. To more firmly establish stress patterns, formal experiments on listener's perceptions of stress would be needed.

While our results with this initial portion of the database do not as firmly verify hypotheses B and C as we had hoped, those hypotheses are still superior to alternative hypothesis J. For example, for talker WAL, we found that, when excluding NP-V boundaries from among the "major constituent boundaries", 79% of the remaining major boundaries were correctly detected by a three-eighth-tone rise in F_0, while 15% of the detected boundaries were than "false", In contrast, only 46% of <u>all</u> stressed syllables by the same talker are detectable from three-eighth-tone rises in F_0.

Bolinger's hypothesis (K) and corollary K1 seem to be confirmed in the small amount of relevant data included in the first subset of sentences. Nine of the 12 occurrences of quantifier-plus-noun combinations included in the speech of the three talkers were <u>not</u> accompanied by substantial (three-eighth-tone) fall-rise valleys. On the other hand, hypothesis L and corollary L1 are <u>not</u> strongly verified by the data summarized in Figure 2. Only 13 of the 50 occurrences of descriptive-adjective-plus-noun show a substantial (three-eighth-tone) F_0 fall-rise valley between the adjective and noun. In general, only a small fraction of the NP's exhibit internal boundaries, regardless of their internal structures. Further studies, using all 99 sentence structures spoken by all three talkers, plus other subsets of the designed sentences, will help establish just when significant F_0 variations do occur within noun phrases. In particular, any contrasts between determiners (quantifiers, numbers, possessives, and combinations of the same) adjectives, participles, and adverbs within NP's can be studied.

[1] Of the 76 sentences with auxiliaries that were spoken by talker LLL, only seven had <u>any</u> F_0 rises, and those seven involved rises of only one or two eighth tones, with the rises usually appearing to be due to phonetic effects, such as F_0 dips during the /w/ of "will".

Also, when we complete the processing on the sentences of the "flying-planes" paradigm (which was not complete at the time of this writing, due to some processing errors), we will be able to determine something about F_0 cues to certain contrastive syntactic structures.

3.6 The Glottal Stop: Interferrence, or Additional Cue?

The all-sonorant sentences were designed to prevent the F_0 jumps and dips due to obstruents from interfering with the investigation of F_0 variations due to stress and syntax. Unfortunately, one effect that was overlooked was the interjection of glottal stops into connected speech when word-initial vowels occur. Glottal stops cause marked variations in F_0, as shown by the examples in Figure 3.

The most immediate impact of glottal stops occurring in the first subset of sentences was that, for 40 of the 600 word and phrase boundaries at which we wished to study F_0 variations, we could not readily separate the stress and syntax effects from the large F_0 variations due to the glottal stops. Those boundaries accompanied by glottal stops were consequently excluded from the results reported in Section 3.5.

There are several reasons for exploring the characteristics of glottal stops and understanding where they may occur in spoken sentences. For one thing, although Potter, Kopp, and Green (1947, p. 80) claim that "there is little possibility of confusing the glottal stop with any of the other stop sounds" when reading spectrograms, practical work on speech recognition schemes shows that glottal stops may exhibit the brief silence and burst characteristics (i.e., high value of spectral derivative) of articulating (oral) stops. When glottal stops are thus interpreted as oral stops, word matching procedures may select the wrong word for that portion of the utterance. At Sperry Univac, we are currently attempting to distinguish such glottal stops from oral stops, using voice onset time, spectral derivatives, and a very low frequency energy function. If F_0 variations can provide additional cues to the occurrence of glottal stops, this would be one instance where F_0 could aid phonetic analysis and word matching.

Another claim about glottal stops is that, while they generally are not phonemic (i.e., word- or meaning-distinguishing) in English (cf. Potter, Kopp, and Green, 1947, p. 80; Smalley, 1963, p. 102), they are "recognized as an indication of stress" (Potter, Kopo, and Green, 1947, p. 80). An examination of the 40 word-initial glottal stops that occurred in the 159 sentences showed that 32 were preceding stressed vowels, while only

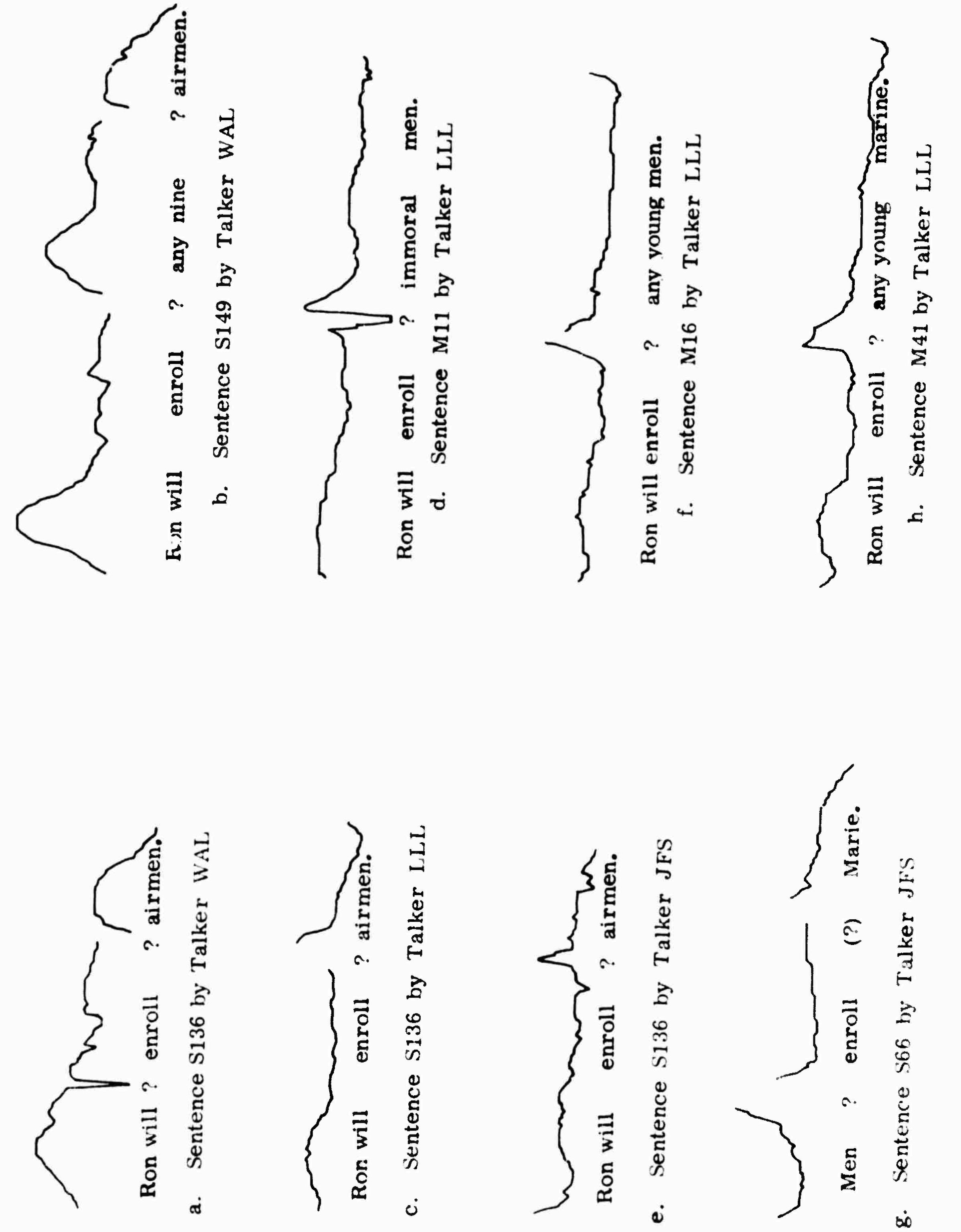


Figure 3. Examples of F_o Contours Near Glottal Stops

eight preceded unstressed vowels. This was true despite the fact that there were 154 word-initial unstressed vowels and only 55 word-initial stressed vowels, before which glottal stops might be expected. Thus, glottal stops are much more likely to precede stressed vowels than unstressed ones, with 57% of the stressed word-initial vowels preceded by glottal stops, while only 5% of the unstressed word-initial vowels were preceded by glottal stops. Clearly, a glottal stop is a likely "indication of stress".

There is also a systematic association of glottal stops with constituent boundaries. While 71% of the stressed word-initial vowels were preceded by glottal stops when they were at major constituent (NP-V and V-NP) boundaries, only 38% of the stressed word-initial vowels within the NP constituent were accompanied by glottal stops. There is a high probability that, given a glottal stop, it occurs at a major constituent boundary. The V-NP boundary, which was found in section 3.5 to be frequently marked by F_0 valleys, is also among the most likely boundaries to be accompanied by a glottal stop.

Glottal stops were first noticed in our analyses by the large F_0 variations they create, as illustrated in Figure 3 (page 27). Part of this typical F_0 change was noted earlier by Potter, Kopp, and Green (1947, p. 80) when they observed that: "The vocal vibration which follows a glottal stop does not always begin at its normal vibratory rate, and an increase in rate of vibration (immediately following the glottal stop) may be observed in the decreasing distance between adjacent vertical striations" on the spectrogram. The slow vibrations of the vocal cords following a word-initial glottal stop have been called "vocal flaps". Some linguists suggest that glottal stops are sometimes achieved by skipping a glottal pulse, thus yielding a brief drop in F_0 to one-half its usual value (T. R. Hofmann, personal communication).

Figure 3 shows both increases in F_0 and decreases in F_0 following glottal stops. Perhaps the simplest characterization would be to say that glottal stops are followed by pronounced variations in F_0 (up or down). Glottal stops also often (though not always) exhibit regions of unvoicing, as in Figures 3a, b, c, d, e, f, and g.

Another aspect of the F_0 contours accompanying glottal stops is the usual increase in F_0 just prior to the glottal stop. This rising pitch preceding glottal stops has been asserted to be a primary source of rising tones in tone languages like Vietnamese (Haudricourt, 1954) and Jinghpaw (Maran, 1971) (cf. also Matisoff, 1973; Hyman, 1973).

The vital role such a rising pitch plays in language change suggests that this F_0 rise is a primary correlate of the glottal stop, physiologically determined, and almost invariably expected to be in the acoustic data. This is the first time it has been observed in Sperry Univac's studies, but it is reported to be a universal phonetic phenomenon (Matisoff, 1973, p. 75).

If so, we might suppose that a procedure for automatically distinguishing between glottal and oral stops would profit from using this F_0 cue. Oral stops are not preceded by such F_0 rises; in fact, it is more likely that an oral stop will show a dip in F_0 near its beginning, even if it is an unvoiced stop (Lea, 1972, 1973b).

We thus see that monitoring the F_0 variations near glottal stops may provide useful information to the acoustic phonetic (i.e., segmentation and labelling) components of a speech understanding system, as well as providing a further cue to the likely occurrence of stress and constituent boundaries.

4. CONCLUSIONS AND FURTHER STUDIES

4.1 Conclusions

Major strides have been made this year in Sperry Univac's study of how to use prosodies to aid speech understanding. An improved computer program for detecting syntactic boundaries (and assigning confidences to the detections) has been developed and delivered to ARPA contractors. Our archetype algorithm for locating stressed syllables has (at last!) been implemented as a computer program and shown to automatically locate 89% of the perceived stresses in connected speech. Both programs have been integrated into speech analysis systems at Sperry Univac and BBN. We are on the threshold of studies to determine how such boundary information may be used to aid syntactic parsing.

Our previous experiments showed a number of important facts about prosodic structures, including these:

- various automatic phonetic labelling procedures work best within stressed syllables
- listener's perceptions provide reliable information about which syllables are stressed.
- the "archetype contour algorithm", which combines F_0, energy, and duration correlates of stress, successfully locates a large majority of stressed syllables and does so better than algorithms based on F_0 alone, or durations of nuclei alone
- F_0 valleys accompany most boundaries between major syntactic constituents
- stressed syllables tend to be roughly equally spaced in time, and pauses at clause and sentence boundaries tend to be integral multiples of the average interstress interval.

This year, we have extended our experimental studies to include some studies of timing cues to linguistic structure, and some initial experiments on the placement of F_0-detected boundaries between major syntactic constituents. A major outcome is that we now know of several potentially useful cues to boundaries between major phrases:

- F_0 valleys
- Lengthened vowels and sonorant consonants
- Long intervals between onsets of stressed syllables
- Pauses at clause and phrase boundaries
- Occurrences of glottal stops.

With the design and recording of our large 3300-sentence data base, we have set the groundwork for extensive controlled experiments on prosodic structure.

We now have considerable evidence that F_0 valleys occur just before the first stressed syllable in major syntactic constituents. While considerable further work is needed to determine exactly which "major syntactic constituents" are marked by such F_0 valleys, it appears that boundaries between main verbs (V's) and following objects of the verb (NP's) are among the most prominently marked sentence divisions, and NP-AUX boundaries are among the least likely to be so marked. Boundaries are not usually marked (by F_0 valleys) within noun phrases, but there are enough instances of such internal boundary markers that further experiments are needed with various NP structures. In any case, we are encouraged that our studies with the initial subset of the designed data base are permitting the controlled testing of various hypotheses relating prosodic patterns to linguistic structures.

4.2 More Experiments with the Data Base

In the near future, we will finish our processing and analysis of the first 99 sentences as spoken by all three talkers. We will continue to study the F_0 contours in the noun phrases with various internal structures, and will explore the regularities concerning the placement of F_0 valleys and glottal stops. We will investigate contrasts in F_0 contours for the contrastive structures of the "flying planes paradigm". When our energy filters become operational, we will also obtain sonorant energy contours, syllabification, and stress location results for those sentences.

We must then continue our study of stress movement and boundary positions with other data base sentences which provide explicit tests of adverbs in various positions, pronouns, expanded auxiliary verbs (including with emphasis and negation), and long complex NP subjects. Simple sentences with identical structures but contrasting phonetic sequences will be studied, to determine more about phonetic influences on F_0 contours and other prosodic patterns.

Following this first set of experiments on stress movement and boundary detections for various syntactic structures, we will then investigate prosodic cues to sentence type. BBN has recently requested that we assign high priority to such studies. This will involve analyzing a subset of sentences which includes declaratives, simple commands, yes/no questions, and WH questions, plus some more complex examples (of each sentence type) that deal with stress movement in the constituents of such sentences.

Prosodic cues to subordination and bracketing are the topic of our other planned experiments. These involve sentences with subordinate clauses, potentially ambiguous placement of clause boundaries, NP-PP-PP sequences, participles, verb-plus-particle constructions, and grouping of conjuncts.

Wherever possible, we will also be studying the timing cues to linguistic structure (such as interstress intervals) for these subsets of data base sentences. We also plan to cooperate with BBN and SDC in the prosodic analysis of some of their data base sentences and protocols.

4.3 More Computer Programs for Prosodic Analysis

The studies of prosodic patterns in the designed sentences and man-computer protocols will probably indicate ways in which the boundary detection program and the stressed syllable location program can be improved. Also, some possible improvements have already been outlined in the documents which described the original versions distributed to ARPA contractors (cf. Lea and Kloker, 1975). The experience that other ARPA contractors will have with tne programs also will probably suggest needed refinements. Based on these continuing studies and applications, we intend to improve the performance, modularity, and clarity of the stress and boundary programs, to increase their utility to speech understanding systems. The restructuring of the prosodic analysis procedures, as shown in Figure 1 of this report (page 6), is one of the first steps to be taken in refining the current implementations at BBN.

We also are considering developing a RHYTHM program, for specifying the time intervals between stress, the number of syllables per second, the occurrence of long disjunctures marking phrase boundaries, and assessments of the <u>rate</u> of speech that may be suitable for aiding phonological analyses.

4.4 Plans for Aiding SUR System Builders

Recently, Sperry Univac and BBN have begun studies of how to use boundary information in the BBN parser. To facilitate such work, BBN has made copies of the BOUND3 and STRESS programs available on System D, so that program refinements can be implemented and tested by Sperry Univac over the ARPANET. The speech data used for such studies will be those travel-budget-task sentences which BBN has incorporated into its speech data base (currently including over 70 snetences). Sperry Univac will study the boundary locations as found by the updated system D version of BOUND3,

and compare them with traces of the parses of sentences obtained from the BBN parser, to see places where prosodic information may have aided the selection of parsing paths. BBN will modify their knowledge source arrays to permit syntax to access any prosodic data that Sperry Univac finds to be helpful.

Issues of continued interest in the BBN-Sperry Univac ineractions include exactly which constituents are demarcated by the boundary detection program and where the detected boundary is located in comparison with the actual time between the two syntactic constituents. It appears from our analyses to date that the detected boundary is usually located after the underlying syntactic boundary, at the position of the last obstruent before the first stressed vowel in the following constituent. Both the studies with the BBN sentences and the controlled experiments described in Sections 3.4 to 3.6 will help determine how to best turn boundary detections into boundary locations suitable for use in parsing procedures.

The Systems Development Corporation (SDC) has also been working on incorporating a version of the boundary detection program into their speech understanding system. The SRI grammar includes specific places where prosodic information can be used to guide syntactic analysis. Sperry Univac plans to interact with SDC (over the ARPANET and by on-site visits) to integrate the prosodic programs into their systems, to try using stress locations to aid word matching, and to investigate using boundaries and stress patterns to aid parsing procedures.

5. REFERENCES

ALLEN, J. (1973), Prosodic Contours for Auxiliary Phrases, Presented at 86th Meeting, Acoustical Society of America, Los Angeles, Calif., November, 1973. Abstract in J. Acoust. Soc. America, vol. 55, No. 1, Jan., 1974.

ALLEN, J. and O'SHAUGHNESSY, D. (1974), Fundamental Frequency Contours of Auxiliary Phrases in English, J. Acoust. Soc. America, vol. 56, Suppl. Fall, 1974, S32(A).

ANDERSON, B.F. (1966). The Psychology Experiment, Belmont, Calif.: Brooks/Cole Publishing Co.

BOLINGER, D. L. (1965), Forms of English: Accent, Morpheme, Order. Cambridge: Harvard Univ. Press.

CHEUNG, J. Y. (1974), Estimating the Magnitude of Linguistic Stress, J. Acoust. Soc. America, vol. 56, S32(A).

CHEUNG, J. Y., HOLDEN, A.D.C., MINIFIE, F.D. (1974), Computer Recognition of Linguistic Stress Patterns in Connected Speech, Proc. IEEE Symposium on Speech Recognition, Carnegie-Mellon University, Pittsburgh, Pa., 142(A).

CHEUNG, J. Y., and MINIFIE, F.D. (1975), Describing Linguistic Stress Contours of Different Sentence Structures, J. Acoust. Soc. Amer., vol. 57, Suppl., Spring, 1975, S25(A).

CHOMSKY, N. (1957), Syntactic Structures. The Hague: Mouton and Co.

CHOMSKY, N. and HALLE, M. (1968), The Sound Pattern of English. New York: Harper and Row.

GILLMAN, R. A. (1975), A Fast Frequency Domain Pitch Algorithm, J. Acoust. Soc. Amer., vol. 58, Suppl. Fall, 1975, S62(A).

HAUDRICOURT, A. G. (1954), De l'Origine des Tons en Vietnamien, Journal Asiatique, vol. 242, 68-82.

HYMAN, L. M. (1973), The Role of Consonant Types in Natural Tonal Assimilations, In Consonant Types and Tone (L. Hyman, ed.). Los Angeles: U. Southern Calif. Press, 151-179.

KLOKER, D. R. (1975a), Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries, presented at the 89th Meeting, Acoustical Society of America, Austin, Texas, April 8-11, 1975. Abstract in J. Acoust. Soc. America, vol. 57, Supp.

KLOKER, D.R. (1975), Phonetic Analysis and Word Matching for Speech Recognition, Univac Report No. PX 11223, Univac Park, St. Paul, Mn.

KURATH, H. (1964), A Phonology and Prosody of Modern English, Ann Arbor: Univ. of Michigan Press.

LEA, W. A. (1972), Intonational Cues to the Constituent Structure and Phonemics of Spoken English, Ph.D. Thesis, School of E.E., Purdue Univ.

LEA, W.A. (1973a), Syntactic Boundaries and Stress Patterns in Spoken English Texts, Univac Report No. PX 10146, Univac Park, St. Paul, Minnesota.

LEA, W.A. (1973b), Segmental and Suprasegmental Influences on Fundamental Frequency Contours. In Consonant Types and Tone (L. Hyman, Ed.), Los Angeles: Univ. of Southern California Press, 15-70.

LEA, W.A. (1973b), An Approach to Syntactic Recognition Without Phonemics, IEEE Transactions on Audio and Electroacoustics, vol. AU-21, 249-258.

LEA, W.A. (1974a), Prosodic Aids to Speech Recognition IV. A General Strategy for Prosodically-Guided Speech Understanding, Univac Report No. PX 10791, Univac Park, St. Paul, Minnesota.

LEA, W.A. (1974b), Sentences for Controlled Testing of Acoustic Phonetic Components of Speech Understanding Systems, Univac Report No. PX 10952, Univac Park, St. Paul, Minnesota.

LEA, W.A. (1975), Isochrony and Disjuncture as Aids to Phonological and Syntactic Analysis, presented at the 89th Meeting of the Acoustical Society, Austin, Texas, April, 1975. Abstract in J. Acous. Soc. of America, vol. 57, Suppl. 1.

LEA, W.A. (To Appear), Sentences for Controlled Testing of Prosodic and Syntactic Components of Speech Understanding Systems, Univac Report No. PX 10953, Univac Park, St. Paul, Minnesota.

LEA, W.A., and KLOKER, D.R. (1975), Prosodic Aids to Speech Recognition: VI. Timing Cues to Linguistic Structure and Improved Computer Programs for Prosodic Analysis, Univac Report No. PX 1123?, Univac Park, St. Paul, Minnesota.

MAEDA, S. (1974), Characterization of Fundamental-Frequency Contours of Speech, J. Acoust. Soc. America, vol. 56, Suppl., S33(A).

MARAN, L. (1971), Burmese and Jinghpaw: A Study of Tonal Linguistic Processes, Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, vol. IV. Also, "A Note on the Development of Tonal Systems in Tibeto-Burman," Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, Vol. II, Urbana, Illinois.

MATISOFF, J. A. (1973), Tonogenesis in Southeast Asia. In Consonent Types and Tone (L. Hyman, ed.), Los Angeles, Calif.: U. Southern Calif. Press, 71-95.

MINIFIE, F.D. and CHEUNG, J.Y. (1975), Measuring Linguistic Stress in a Continuum, J. Acoust. Soc. America, vol. 57, Suppl., S25(A).

POTTER, R.K., KOPP, G.A., and GREEN, H.C. (1947), Visible Speech. New York: D. van Nostrand Co.

SARGENT, D.C., LI, K.P., and FU, K.S. (1974), Some Problems Associated with Automatic Stress Contour Extraction, Proc. IEEE Symposium on Speech Recognition, Carnegie-Mellon University, Pittsburgh, Pa., 142(A).

SKINNER, T.E. (1975), Acoustic-Phonetic Analysis of Speech, Univac Report No. PX 11222, Univac Park, St. Paul, Minnesota.

SMALLEY, W.A. (1963), Manual of Articulatory Phonetics. Ann Arbor, Mich.: Cushing-Malloy, Inc.

TRAGER, G.L. and SMITH, H.L., Jr. (1951), An Outline of English Structure, Studies in Linguistics: Occasional Papers 3, Norman, Oklahoma: Battenburg Press.

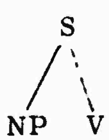
WOODS, W.A., et al. (1975a), Speech Understanding Systems, Quarterly Technical Progress Report No. 2 (BBN Report No. 3080), Volt Beranek and Newman, Inc., Cambridge, MA.

WOODS, W.A., et al. (1975b), Speech Understanding Systems, Quarterly Technical Progress Report No. 3 (BBN Report No. 3115), Bolt Beranek and Newman, Inc., Cambridge, MA.

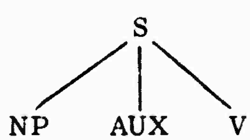
6. APPENDIX:
SENTENCES FOR TESTING BOUNDARY PLACEMENTS

The sentences selected for the first subset of the data base are listed in this Appendix. The full set of 99 sentences was analyzed for one talker (LLL) and the first 37 of those sentences (i.e., subsets 1A and 1B) were analyzed for two other talkers (WAL and JFS). Each sentence is given a designator, such as "PSS S2" or "PSS M27", where PSS stands for "prososyntactic sentence", "S" stands for short and "M" for medium length, and the number indicates the order of that sentence in the original systematic ordering of the four equal divisions of the data base into short, medium, long, and obstruent types (Lea, To Appear; cf. also Section 3.4).

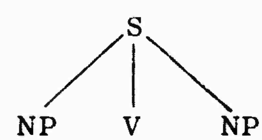
Also accompanying each sentence is an indication of the expected stress pattern and bracketing of the sentence (e.g., S[SU][US] indicates a stressed monosyllabic constituent, followed by a constituent with a stress then an unstressed syllable, followed by a third constituent consisting of an unstressed and a stressed syllable). The T plus number indicates which syntactic tree that sentence exhibits, based on the following six trees:



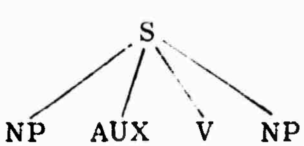
(T1)



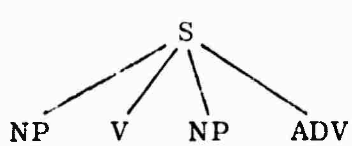
(T2)



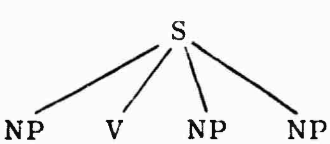
(T3)



(T4)



(T5)



(T6)

SUBSET 1A. One Stress Per Constituent.

PSS	S2 – SS,T1	Men know.
PSS	S4 - SUS, T2	Men will know.
PSS	S8 - SSS, T3	Men know Ron.
PSS	S12 - SUSS, T4	Men will know Ron.
PSS	S16 - SSSS, T5	Men know Ron now.
PSS	S24 - SSSS, T6	Men owe Ron rum.
PSS	S26 - S(US), T1	Men enroll.
PSS	S28 - S (SU), T1	Men worry.
PSS	S37 - SU(US), T2	Men will enroll.
PSS	S39 - SU(SU), T2	Men will worry.
PSS	S42 - SS(US), T3	Men know Marie.
PSS	S43 - SS (SU), T3	Men know Mary.
PSS	S47 - S (US) S, T3	Men enroll Ron.
PSS	S51 - S (SU) S, T3	Men worry Ron.
PSS	S56 - SS(UUSU), T3	Men know Leonora.
PSS	S57 - SS(USU), T3	Men know Maria.
PSS	S58 - SS(SUU), T3	Men know Melanie.
PSS	S66 - S (US) (US), T3	Men enroll Marie.
PSS	S67 - S (US) (SU), T3	Men enroll Mary.
PSS	S68 - S(SU) (US), T3	Men worry Marie.
PSS	S69 - S(SU) (SU), T3	Men worry Mary.
PSS	S107 - S S(UUS), T3	Ron knew a marine.
PSS	S108 - SS (USU), T3	Ron knew an airman.
PSS	S136 - S U(US)(SU), T4	Ron will enroll airmen.

SUBSET 1B. Two or More Stresses Per Constituent: Expansions of Determiner.

PSS	S137-S U(US)(SS), T4	Ron will enroll nine men.
PSS	S138-S U(US)(SS)*, T4	Ron will enroll your men.
PSS	S139-S U(US) (SS), T4	Ron will enroll all men.
PSS	S140-S U(US) (SS), T4	Ron will enroll no men.
PSS	S141-S U(US) (SUS), T4	Ron will enroll any men.
PSS	S142-S U(US) (SUS), T4	Ron will enroll many men.
PSS	S143-S U(US) (SUS), T4	Ron will enroll nine airmen.
PSS	S144-S U(US) (SSU)*, T4	Ron will enroll your airmen.
PSS	S145-S U(US) (SSS)*, T4	Ron will enroll all your men.
PSS	S146-S U(US) (SSS), T4	Ron will enroll all nine men.
PSS	S147-S U(US) (SUSS), T4	Ron will enroll any nine men.
PSS	S148-S U(US) (SSU), T4	Ron will enroll all nine airmen.
PSS	S149-S U(US) (SUSSU), T4	Ron will enroll any nine airmen.

*The initial prediction was that "your" would be stressed. It now seems more likely that "your" will be unstressed in these sentences.

SUBSET 1C. Prenominal Adjectives, Participles, and Adverbs (with 4 or less syllables in NP)

PSS	M1 - S S (US) (SS), T4	Ron will enroll young men.
PSS	M2 - S S (US) (SS), T4	Ron will enroll real men.
PSS	M3 - S S (US) (SUS), T4	Ron will enroll moral men.
PSS	M4 - S S (US) (SUS), T4	Ron will enroll willing men.
PSS	M5 - S S (US) (SSU), T4	Ron will enroll young airmen.
PSS	M6 - S S (US) (USS), T4	Ron will enroll a young man.
PSS	M7 - S S (US) (SSS), T4	Ron will enroll all young men.
PSS	M8 - S S (US) (SSS), T4	Ron will enroll nine young men.
PSS	M9 - S S (US) (SSS)*, T4	Ron will enroll your young men.
PSS	M10 - S S (US) (SSS), T4	Ron will enroll mean young men.
PSS	M11 - S S (US) (USUS), T4	Ron will enroll immoral men.
PSS	M12 - S S (US) (USUS), T4	Ron will enroll a moral man.
PSS	M13 - S S (US) (USUS), T4	Ron will enroll a young marine.
PSS	M14 - S S (US) (SUUS), T4	Ron will enroll mannerly men.
PSS	M15 - S S (US) (USSU), T4	Ron will enroll a young airman.
PSS	M16 - S S (US) (SUSS), T4	Ron will enroll any young men.
PSS	M17 - S S (US) (SUSS), T4	Ron will enroll many young men.
PSS	M18 - S S (US) (SSUS), T4	Ron will enroll only young men.
PSS	M19 - S S (US) (USSS), T4	Ron will enroll nine moral men.
PSS	M20 - S S (US) (USSS), T4	Ron will enroll a new young man.
PSS	M21 - S S (US) (USSS), T4	Ron will enroll a mean young man.
PSS	M22 - S S (US) (SSSU), T4	Ron will enroll nine young airmen.
PSS	M23 - S S (US) (SSSU) *, T4	Ron will enroll your young airmen.
PSS	M24 - S S (US) (SSSS) *, T4	Ron will enroll all your young men.
PSS	M25 - S S (US) (SSSS)*, T4	Ron will enroll your nine young men.
PSS	M26 - S S (US) (SSSS) *, T4	Ron will enroll your new young men.
PSS	M27 - S S (US) (SSSS), T4	Ron will enroll new mean young men.

*The initial prediction was that "your" would be stressed. It now seems more likely that "your" will be unstressed in these sentences.

SUBSET 1D. Prenominal Adjectives, Participle, and Adverbs (with more than 4 syllables in NP)

PSS	M28 - S U (US) (USUUS), T4	Ron will enroll a moral marine.
PSS	M29 - S U (US) (UUSUS), T4	Ron will enroll an immoral man.
PSS	M30 - S U (US) (USUSU), T4	Ron will enroll immoral airmen.
PSS	M31 - S U (US) (USUSU), T4	Ron will enroll a moral airman.
PSS	M32 - S U (US) (USUSU), T4	Ron will enroll a lonely airman.
PSS	M33 - S U (US) (USUSS), T4	Ron will enroll a moral young man.
PSS	M34 - S U (US) (USSUS), T4	Ron will enroll a young moral man.
PSS	M35 - S U (US) (SUSUS), T4	Ron will enroll moral lonely men.
PSS	M36 - S U (US) (SUSUS), T4	Ron will enroll lonely moral men.
PSS	M37 - S U (US) (SUSUS), T4	Ron will enroll many moral men.
PSS	M38 - S U (US) (SUSUS), T4	Ron will enroll only moral men.
PSS	M39 - S U (US) (USUUS), T4	Ron will enroll many willing men.
PSS	M40 - S U (US) (SUSUS), T4	Ron will enroll nine immoral men.
PSS	M41 - S U (US) (SUSUS), T4	Ron will enroll any young marine.
PSS	M42 - S U (US) (SSUSU), T4	Ron will enroll nine moral airmen.
PSS	M43 - S U (US) (SUSSS), T4	Ron will enroll lonely mean young men.
PSS	M44 - S U (US) (SSUSS), T4	Ron will enroll new moral young men.
PSS	M45 - S U (US) (SSSUS), T4	Ron will enroll new young moral men.
PSS	M46 - S U (US) (SSS), T4	Ron will enroll well known men.
PSS	M47 - S U (US) (SUSS), T4	Ron will enroll really young men.
PSS	M48 - S U (US) (SUUSUS), T4	Ron will enroll really immoral men.
PSS	M49 - S U (US) (SSUUSS), T4	Ron will enroll really mannerly young men.
PSS	M50 - SU (US) (SUSSS), T4	Ron will enroll really well known men.
PSS	M51 - SU (US) (SUSUSUSS), T4	Ron will enroll really immoral well known men.
PSS	M52 - SU (US) (SUSSUSUS), T4	Ron will enroll really well known immoral men.
PSS	M53 - SU (US) (SUSUUSUS), T4	Ron will enroll really willing immoral men.
PSS	M54 - SU (US) (SUSUUSUSS), T4	Ron will enroll really willing immoral young men.

SUBSET 1E. "Flying-Planes" Paradigm

PSS	M229 -(US)	U	(SUS), T3	Lawmen are lying men.
PSS	M230 - (SU)	U	(SU) S, T4	Lawmen are ruling Maine.
PSS	M231 - (US)	U	(SUS), T3	Airmen are lying men.
PSS	M232 - (SU)	U	(SU) (SU), T4	Airmen are eyeing women.
PSS	M233 - (SU)	U	(SUS), T3	Airmen are erring men.
PSS	M234 - (SU)	U	(SU) S, T4	Women are airing wool.
PSS	M235 - (SU)	U	(SUS), T3	Lawmen are rummy men.
PSS	M236- (SU)	U	(SU) (SU), T4	Lawmen are ruling women.